2023 IEEE 26th International Conference on
Intelligent Transportation Systems (ITSC)
24-28 September 2023. Bilbao, Bizkaia, Spain

A Multi-Sensor Video/LiDAR System for Analyzing Intersection Safety

Aotian Wu, Tania Banerjee, Ke Chen, Anand Rangarajan and Sanjay Ranka

***Abstract*—We introduce an integrated video and LiDAR analytics system for analyzing pedestrian and vehicle behavior at traffic intersections. Subsystems for each modality leverage advanced deep-learning techniques to detect pedestrians and vehicles and then use a Kalman-filter-based tracking algorithm to generate tracks. The video and LiDAR tracks are then aligned spatiotemporally onto the same coordinate system with synchronized clocks.**

We evaluate the benefits of these two modalities by providing both qualitative and quantitative comparisons, utilizing low-level measures such as detection and tracking accuracy, as well as high-level measures such as severe events. Additionally, we compare the two modalities at different times of the day and show that LiDAR is competitive with video during daylight hours and significantly outperforms video at late evening when lighting conditions are poor. To the best of our knowledge, this study represents the first detailed comparison of these two modalities for observing traffic intersections.

***Index Terms*—Pedestrian Safety, Surrogate Measures, Near-misses, Severe Events**

I. Introduction

Intersections are often considered high-risk areas for traffic crashes, as they are points of conflict between vehicles traveling in different directions. Nearly 28% of all fatal crashes and 58% nonfatal crashes are intersection crashes, resulting in $179 billion, or 53% of all economic costs from motor vehicle crashes [1]. Toward the goal of improving intersection safety, it is crucial to build a traffic monitoring and analytics system that helps understand road user behaviors, identify abnormal behaviors and patterns, assess potential risks, and prevent crash events proactively. Many intersections are already equipped with surveillance cameras, making it easier to integrate video monitoring into the existing infrastructure. LiDAR has also gained attention due to its reduced cost and increased precision, making it a potential alternative or complement to camera-based systems for intersection monitoring.

Camera-based systems are generally adopted for intersection monitoring and analysis [2], [3] due to their low cost and advanced object detection and tracking algorithms. Video cameras provide a rich visual representation of the scene, which can be useful for understanding traffic patterns and behaviors. However, they may be limited for the following scenarios: (1) poor lighting and weather conditions; (2) occlusions caused by infrastructure (e.g., poles) or other road users; (3) pedestrian detection in areas where image footprint is small; and (4) areas where the field of view (FOV) does not cover. Additionally, the lack of 3D understanding of the scene and distortion in case of fisheye cameras may result in an inaccurate estimate of speed and severity of conflict events.

Aotian Wu, Tania Banerjee, Ke Chen, Anand Rangarajan, Sanjay Ranka are with the University of Florida, Gainesville, FL 32611.

LiDAR provides a highly accurate 3-D representation of the surrounding environment in terms of point clouds because it precisely measures the 3-D distance and scale of objects by measuring the time between emitting and receiving the reflected laser light waves. Additionally, it has long-range detection capabilities and is robust under different lighting conditions. Thus, it has the potential to be complementary to vision sensors. LiDAR has its own potential limitations, especially considering the lack of appearance information and occlusion issues. Previous studies [4]–[6] have shown the effectiveness of LiDAR in infrastructure analytic systems. They produce trajectories in four main steps: (1) background filtering, (2) object clustering, (3) object classification, and (4) object tracking. However, there are many challenges that need to be addressed. For example, both background filtering and object clustering require extensive parameter tuning to achieve satisfactory results. Moreover, in object clustering, it is common for one object to be assigned to multiple clusters, or for multiple nearby smaller objects to be assigned to the same cluster, resulting in tracking challenges. One study [7] leverages deep convolutional neural networks pretrained on autonomous driving datasets, but only achieves suboptimal detection performance (72.9% average F1 score).

In this paper, we present an integrated video and LiDAR analytics system for analysing pedestrian and vehicle behavior at traffic intersections. Subsystems for each modality leverage advanced deep-learning techniques to detect road users and then use a Kalman-filter-based tracking algorithm to generate tracks. The key contributions of the paper are as follows:

1) Spatiotemporal mapping of the video and LiDAR tracks. This involves mapping the two modalities to the same space, namely the Google Maps coordinate system, and clustering them based on their ingress and egress lanes. Additionally, we develop simple clock synchronization techniques to account for small timing differences in data acquisition between the two sensors.
2) Using this system, we evaluate the benefits of the two modalities and provide both a qualitative and quantitative comparison. Utilizing both low level measures such as detection and tracking accuracy and high level measures such as severe events [8], we provide a qualitative and quantitative comparison of the two modalities during different times of day. Our definition of severe events is a generalization of near-misses and incorporates speed

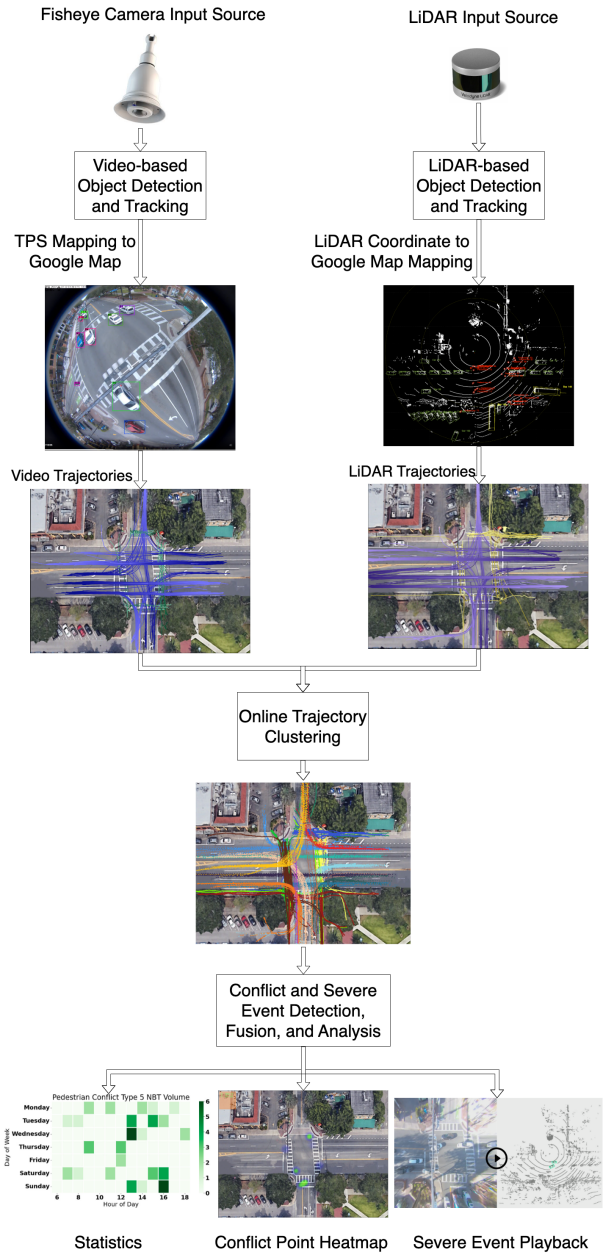


Fig. 1. The overall workflow of our system. From an input video and LiDAR stream, the system initially detects and tracks all road users for each of the two streams, then maps them to the same Google Maps coordinate system. The trajectories are then clustered based on their movement patterns. Finally, the conflict and severe event detection and analysis module fuses this information to generate statistics for severe events, heatmaps, and video playback clips.

and acceleration profiles, clustering information, and the direction of the trajectories [8], and is used for both vehicle-vehicle and pedestrian-vehicle interactions.

3) We develop techniques for fusing severe conflict events to produce statistics, a conflict point heatmap, and video footage of severe events in which the LiDAR and video are synchronized and displayed side by side.

Experimental results are presented using video and LiDAR data collected concomitantly at the same intersection. Our overall results show that LiDAR is competitive with video during the day and significantly outperforms video during early morning and late evening when the lighting conditions are poor. We currently provide users with a union of the severe events found by two modalities. In the future, we plan to intelligently combine the two modalities by automatically choosing the better of the two based on location of the intersection (e.g., areas where LiDAR is more accurate than video and vice versa), traffic conditions (e.g., when video accuracy is reduced due to occlusions), and lighting and weather conditions.

The rest of the paper is structured as follows. The methodology for developing the integrated system is described in Section II. Section III outlines the experiments conducted on an intersection in Gainesville, Florida, to demonstrate the advantages of our system. Section IV presents the related work on the use of cameras and LiDAR sensors for monitoring intersections. Conclusions are described in Section V.

II. Methodology

The overall workflow of the video and LiDAR analytic system is shown in figure 1. For an input video and LiDAR stream, the system first detects and tracks all road users, then maps to the same Google Map coordinate system. At this stage, the trajectories are clustered based on their movement patterns. Finally, the conflict and severe event detection and analysis module generates severe event statistics, heatmaps, as well as video playback clips.

A. Video-based Detection, Tracking, and Mapping

The video-based object detection and tracking module utilizes YOLOv4 [9] to detect different kinds of road participants, including vehicles, pedestrians, cyclists, and motorcyclists. A modified version of the DeepSORT [10] algorithm is used to associate detections across frames and assign a unique ID for each object. The modification is necessary because of the large distortion in fisheye videos. Specifically, the trajectories in fisheye videos have unusual shapes and speeds, which do not work well with the Kalman Filter used in DeepSORT. Therefore, instead of computing the distances in the original fisheye coordinates, we first align them to Google Map coordinates and then compute distances in this regular coordinates. Details of the mapping and overall approach can be found in [11].

B. LiDAR-based Detection, Tracking, and Mapping

We used state-of-the-art algorithms to detect and track objects in LiDAR point clouds, specifically Centerpoint [12] and SimpleTrack [13]. The detector was trained on high-quality annotations using an efficient annotation tool [14] and can detect traffic participants with high precision.

LiDAR-based Object Detection: We detect road users using CenterPoint [12], which identifies objects as key points and regresses their other attributes, such as 3-D location, size, and heading orientation. CenterPoint consists of a standard 3-D backbone network, a center heatmap head, and regression

heads. The center heatmap head produces keypoint heatmaps, where each peak corresponds to a predicted object center. The regression heads regress other properties for predicted key points. We followed OpenPCDet's [15] implementation of CenterPoint. More details on the model can be found in [12].

LiDAR-based Object Tracking: We used SimpleTrack [13] for multi-object tracking. SimpleTrack is a high-performing multi-object tracking approach that unifies 3D MOT methods into a general framework. It consists of four major components: detection preprocessing, object motion modeling, BBox association across frames, and tracklet lifecycle management. We apply stricter non-maximum suppression (NMS) to exclude overlapping low-confidence BBoxes while preserving low-confidence BBoxes caused by sparsity or occlusion. For object motion modeling, we use the Kalman filtering algorithm, which estimates the next locations of objects from uncertain observations. We found that the Kalman filter performs well in infrastructure-based LiDAR settings due to the high-precision measurements of objects subject to kinematics. We associate the predicted location with detections in the next frame using the Hungarian algorithm [16].

Map the 3D Trajectories to Google Maps Coordinates: For each 3D point, we map it to a 2D point on Google Maps using the following equation:

$$\begin{bmatrix} x_g \\ y_g \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & t_x \\ \sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} \quad (1)$$

where (x_p, y_p, z_p) are centroids of the 3D BBoxes, (x_g, y_g) are mapped points in Google Maps coordinates, θ is the rotation angle, t_x, t_y are the translation offsets, and s_x, s_y are the scaling parameters along the x and y axes accordingly. Notice that z_p is omitted because the LiDAR is parallel to the ground.

C. Trajectory Clustering

To accurately identify and categorize severe conflicting events, we first need to cluster trajectories based on their movement directions. We represent the movement directions using phases, as illustrated in Figure 2. Certain phases are in conflict with each other, while others are not. For instance, Phase 1 is in conflict with Phase 2 but not with Phase 5, as shown in Figure 2. We are particularly interested in interactions between trajectories in conflicting phases, which may reveal potential risks of the intersection. The trajectories are clustered in two modes: online and offline. Online clustering matches each new trajectory as it happens with the list of representative trajectories returned from offline clustering and assigns a cluster to the new trajectory. The best matching trajectory is picked by computing the Dynamic time warping (DTW) distance.

Offline clustering repeats every 24 hours and results in a representative trajectory (centroid) for each cluster. Specifically, it groups together the trajectories based on ingress and egress lanes. Each possible combination of ingress-egress lanes is considered a distinct class, assuming that vehicles do not change lanes at intersections. Trajectories that do involve lane changes are considered anomalous trajectories. Then, we compute a representative trajectory for each cluster by averaging multiple trajectories of the same cluster. This concludes the offline clustering.

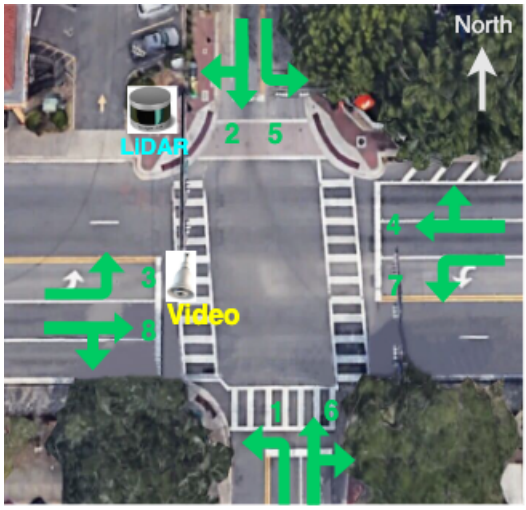


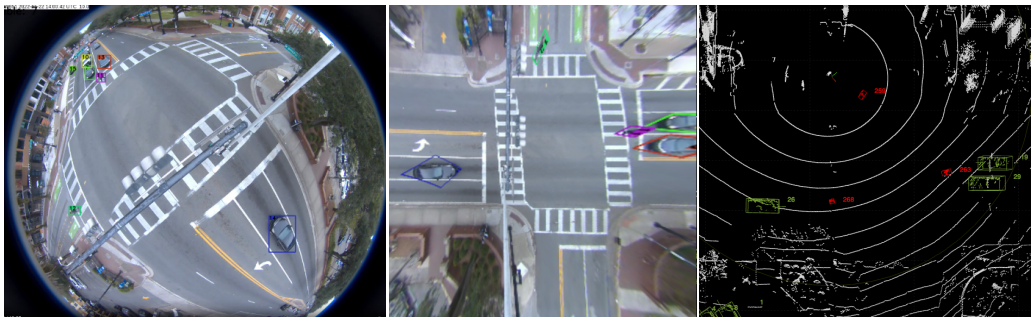
Fig. 2. Different traffic movements at the intersection used for our experimental results. The traffic movements are assigned a number between 1 and 8. Additionally, the positions of both the camera and LiDAR sensor are depicted within the diagram.

D. Severe Events

Using trajectories extracted from video and LiDAR, we can conduct safety analysis to identify potential risk factors at intersections. For each trajectory, we compute its speed and acceleration at each timestamp, along with the lane and cluster belonging information. For pairs of trajectories passing the intersection within the same time period, we compute surrogate safety measures, namely time-to-collision (TTC) and post-encroachment time (PET). TTC measures the time it takes for two road users to collide if they keep their current velocity. PET is the time difference between one road user leaving the conflict zone and the other road user entering the zone. A low TTC or PET value indicates a high risk of collision.

The output of our safety analysis system is a list of near-miss events. To ensure the accuracy of detecting these events, we propose a two-stage filtering approach. In the first stage, we identify potential severe near-miss events using loose criteria: (1) Both road users pass the intersection at the same time; (2) they are in conflicting traffic phases; (3) either TTC or PET is less than 10 seconds; (4) both of them are moving; and (5) the distance between them is less than 10 meters. We refer to the pair of trajectories passing the first-stage events as conflict events in the rest of the paper. Conflict events provide good coverage of potentially dangerous situations.

In the second stage, we employ domain expertise to conduct more fine-grained filtering. We refer to the events passing this second stage as severe events. In this stage, we use the following criteria to further narrow down the conflict events. For vehicle-to-vehicle (V2V) conflicts, we check for two conditions: (1) TTC/PET is less than 3 seconds and both vehicles



(a) (b) (c)

Fig. 3. An example of detection results on video and LiDAR subsystems. (a) The detected bounding boxes on the original fisheye image. (b) The rectified detection result aligned with the Google map coordinates. (c) The LiDAR detection results on the same Google Map coordinates.

are moving relatively fast; and (2) the vehicles properly yield to each other (deceleration is detected). For pedestrian-to-vehicle (P2V) conflicts, we also check (1) whether TTC/PET is less than 3 seconds; (2) whether the pedestrian is moving at a reasonable speed (to exclude the cases where cyclists or motorcycles are wrongly detected as pedestrians); and (3) whether the pedestrian has left the curb.

We will like to note that many of the thresholds (e.g., distance and time) are tunable by the user and in general will depend on the intersection and the city. In our future work, as we collect data from more intersections, we will develop techniques for automatically setting these thresholds.

E. Fusion of LiDAR and Video Data

We conduct the safety analysis separately for video and LiDAR before fusing the output. Fusion is performed at the severe event level. We chose to do this to reduce the overall computational overhead as we are mainly interested in comparing the two modalities using higher level measures.

The initial step is the synchronization of the two sources. We select a LiDAR frame and find a video frame that roughly matches the placement of road users. Then, we consider the 10 surrounding timestamps of that video frame as candidates. For each possible timestamp, it is assumed to align with the LiDAR frame's timestamp, and the average distance between the trajectories of all objects that appeared in both LiDAR and video is computed. The timestamp with minimum average distance is considered the match. This process is repeated for tens of chosen LiDAR frames. Using the average time difference between multiple pairs of LiDAR frames and their matching video frames, we synchronize the two sources.

For V2V severe event fusion, we first find matching video and LiDAR trajectory pairs. In other words, a video severe event matches a LiDAR severe event if the corresponding trajectories match each other (this is based on same timestamp of the two frames and average distance between the two trajectories to be within a small user defined threshold). If the minimum distance exceeds the threshold, it is determined that there is no match. The same process is also used to find a video trajectory that matches a given LIDAR trajectory.

In addition to the strictly matched events, in which both video and LiDAR consider them severe and both trajectories are matched, we also defined partially matched events. A partially matched event indicates that at least one trajectory is matched and at least one sensor considers it severe. If a video or LiDAR event cannot find a strictly or partially matched event from the other source, it is considered a video-only or LiDAR-only event. In the ideal world, the video and LiDAR trajectories should match - but in practice this is not the case due to distortion, occlusion, lighting conditions and processing errors that are variable for the two modalities.

For P2V event fusion, due to the challenges in matching pedestrian trajectories and the possibility of one vehicle conflicting with multiple pedestrians, we have decided to relax the severe event matching criteria. As a result, if the involved vehicles in the video and LiDAR severe events match, even if the pedestrian trajectories do not precisely match, we will consider it a match.

F. Visualization

In addition, we provide event visualization. We extract a 10-second clip surrounding the conflict time for each Severe event. In the clip, we display rectified video frames and a top-down view of the LiDAR frames, highlighting the two or more objects involved in the conflict. Note that for P2V events, we highlight all involved and potentially at-risk pedestrians. The event playback enables traffic engineers to closely examine each event, analyze driver behavior, and identify dangerous maneuvers. Detailed analysis results and visualization examples are provided in Section III-C.

III. Experiments

A. Sensor and System Setup

We collected the datasets at a busy intersection—West University Avenue & Northwest 17th Street, Gainesville, FL—near the campus of the University of Florida. Both sensors are mounted parallel to the ground on traffic posts. As shown in Figure 2, the LiDAR sensor is mounted on the northwest corner of the intersection, while the fisheye camera is mounted in the center of the west side. The camera sensor is a GRIDSMART bell camera with a 180° fisheye lens.

TABLE I
DETECTION PERFORMANCE OF VIDEO AND LIDAR IN DAY TIME AND NIGHT TIME (TP - TRUE POSITIVE, FN - FALSE NEGATIVE, FP - FALSE POSITIVE).

	Day time				Night time			
	Vehicle		Pedestrian		Vehicle		Pedestrian	
	video	LiDAR	video	LiDAR	video	LiDAR	video	LiDAR
TP	4395	9883	485	4141	3063	9993	467	10568
FN	654	52	137	11	1100	32	362	151
FP	228	56	40	334	482	34	97	1537
Recall	87.0	99.5	78.0	99.7	73.6	99.7	56.3	98.6
Precision	95.1	99.4	92.4	92.5	86.4	99.7	82.8	87.3
F-1 score	90.9	99.5	84.6	96.0	79.5	99.7	67.0	92.6

The LiDAR sensor is a Velodyne VLP-32C LiDAR with 32 channels, a 200-meter range, $+15°$ to $-25°$ vertical field of view (FOV), and 360° horizontal FOV. Both sensors have a frame rate of 10 Hz.

Our video and LiDAR detection and tracking framework was executed on the NVIDIA TITAN RTX GPU, which is built on the Turing architecture and features 4608 CUDA cores, 576 Tensor cores, and 72 RT cores. With 24 GB of GDDR6 memory and a memory bandwidth of 672 GB/s, the TITAN RTX is a high-performance computing device that can handle large and complex datasets.

In our implementation, video and LiDAR pipelines share the online trajectory clustering, near-miss, and severe event detection modules, but have separate object detection, tracking, and coordinate mapping modules. The modules are containerized using Docker. Between modules, we utilize RabbitMQ to pass messages with negligible latency. The pipeline is triggered by adding a new video file or LiDAR PCAP file to the specified paths. And the detected near-misses and severe events, as well as the intermediate output such as detection and tracking results, are stored in a MySQL database.

B. Perception Analysis

This section evaluates the accuracy of detecting vehicles and pedestrians in video and LiDAR during the same time periods. We carefully annotate 5-minute daytime and 5-minute nighttime video and LiDAR sequences. The detection result is shown in Table I. It is worth noting that the number of ground truths is different for video and LiDAR because they have different detectable ranges and occlusion areas. An example of the detection result is shown in Figure 3. The majority of their detections correspond, but the pedestrian on the left of the LiDAR frame is occluded in the video.

C. Severe Event fusion analysis

We collected data from both sensors from 7 a.m. to 10 p.m. for one week as well as running them through the pipeline which outputs the severe events.

Vehicle-to-vehicle: In the top portion of Figure 4, we show the count of correctly detected severe events. The "matched" section includes both strictly and partially matched cases. It shows that video and LiDAR are able to detect the majority of

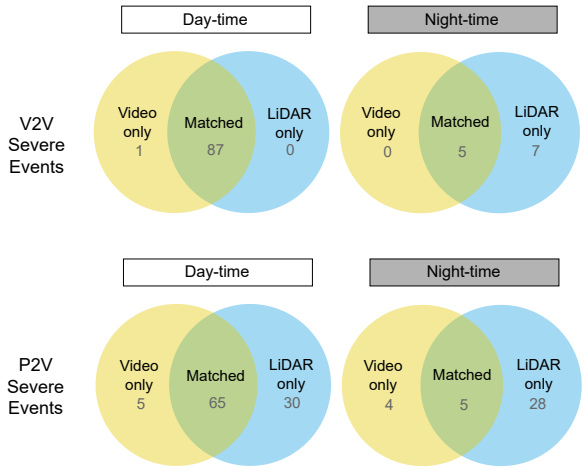


Fig. 4. Statistics of daytime and nighttime V2V and P2V severe events detected by video, LiDAR, or both.

TABLE II
STATISTICS OF MATCHING VIDEO AND LIDAR VEHICLE-TO-VEHICLE SEVERE EVENTS.

	Video	LiDAR
All	84	97
Strictly matched	51	51
Partially matched	19	26
No match	14	20
Strictly matched coverage	60.7%	52.6%
Loosely matched coverage	83.3%	79.4%

severe events during the daytime, while LiDAR detects more severe events at nighttime. Detailed statistics are shown in Table II. Unlike the Venn diagrams, this table also includes false positives. For example, there are $51 + 19 + 26 = 96$ matched events shown in the table, for both partially and strictly matched ones. However, there are only $87 + 5 = 92$ matched events shown in Figure 4. This discrepancy comes from 4 false positive events. These false positives were identified through manual assessment of playback clips for each severe event. We also calculate the matched coverage, where "loosely matched" includes both strictly matched and partially matched events. A high percentage of overlapped events demonstrates the accuracy of both sensors, as both are capable of detecting the same potentially dangerous events. Due to the speed and location imprecision of sensors, partially matched cases may result from only one sensor meeting the severe event criteria, while the other does not. We further analyzed the no-match cases and found the main reasons to be as follows: (1) only one sensor data is available, while the other is either corrupted during that time or incomplete; (2) the video's detection accuracy degrades at night, resulting in both false positives and false negatives; and (3) some objects of minor classes (such as trucks and cyclists) are not detected by LiDAR (This is currently the limitation of our training dataset); (4) Large trucks and buses tend to occlude LiDAR's view compared to video because of different placement positions.

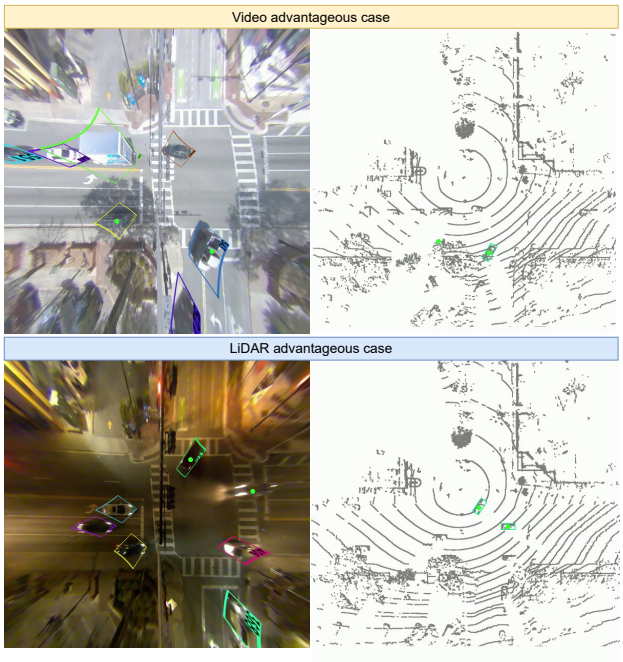


Fig. 5. Example of video and LiDAR snapshot of a vehicle-to-vehicle severe event found using our system. The involved vehicles are highlighted with green dots. A green dot without a corresponding bounding box suggests a false negative.

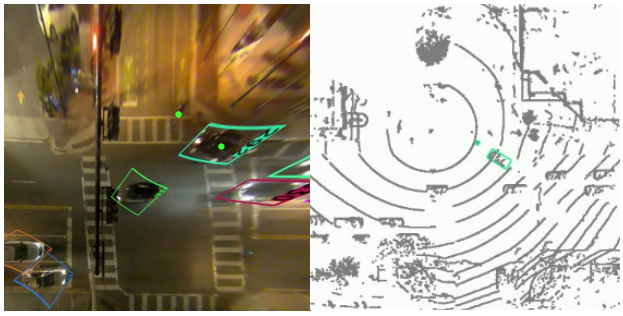


Fig. 6. Example of LiDAR snapshot of a vehicle-to-pedestrian event at night. Only LiDAR reports this event because the video detector fails to detect the pedestrian due to poor lighting conditions.

Pedestrian-to-vehicle: The lower portion of Figure 4 shows the detected P2V severe events. These results show that LiDAR is superior for both daytime and nighttime detection of pedestrian-vehicle severe events. The gap is considerably higher at nighttime. The main reason for its advantage during daytime is mainly due to the difference in detection range. When annotating pedestrians in the video, there were coverage areas on the intersection where the image footprint of pedestrians in the fisheye space was very small. LiDAR is able to detect pedestrians up to 50 meters away. We noticed that LiDAR detects numerous instances of pedestrians crossing vehicle lanes that are not detected by video. For nighttime, most of the P2V severe events are only detected by LiDAR. As shown in Figure 6, pedestrian visibility at night is extremely low for both human and camera-based detectors.

Fig. 7. The heatmaps for pedestrian-to-vehicle and vehicle-to-vehicle severe based on common conflict regions using both LiDAR and video streams using our system.

Figure 6 shows a case in which a right-turning vehicle noticed the pedestrian in the middle of a turning maneuver, placing the pedestrian in a dangerous situation. We believe LiDAR sensors are crucial for improving nighttime pedestrian safety. Although some of the variability can be attributed to the level of annotation and training conducted for each of the two modalities, we believe that our conclusions are relatively independent of this variability.

Finally, instead of case-by-case severe event visualization, we also generate V2V and P2V heatmaps summarizing the most common conflict zones, as shown in Figure 7. The heatmap highlights the regions where traffic engineers should focus their attention to prevent accidents in advance.

D. Computational Speed

The detection, tracking, and mapping modules are implemented in an online fashion, i.e., frame-by-frame data processing and message passing. We measure the running time using a full day of data. The video-based detection and tracking run at 40 Hz (40 frames per second). The LiDAR-based detection runs at 12 Hz and the tracking runs at 10

Hz. The message passing and mapping for both sensors take negligible time. Note that the data acquisition rate for each pipeline is 10 Hz. LiDAR processing is mostly real-time, but it can be slowed during peak traffic times when a large number of objects are present resulting in more trajectories. Video processing is considerably faster than LiDAR. The downstream modules, namely trajectory clustering, conflict, and severe event detection, process input in batches and are currently offline. These computations are limited to the number of objects (rather than the size of the images) and generally are extremely fast.

IV. Related Work

Intersection monitoring using video cameras has been the focus of several studies in recent years. In [17], Cheung and Kamath proposed a system that utilizes video cameras to detect and track vehicles at intersections using a combination of foreground validation and background subtraction. As part of our previous work [18], we developed an integrated two-stream convolutional network architecture for real-time detection, tracking, and near-miss detection. Related work by Qi et al. [19], also proposed a computer-vision-based intelligent system for analyzing traffic at road intersections. Overall, these studies demonstrate the effectiveness of video-based intersection monitoring systems in detecting and tracking vehicles, classifying road users, and analyzing traffic patterns. However, challenges such as lighting and weather conditions, occlusions, and lack of 3D understanding of the scene, can affect the reliability of these systems, and there is still room for improvement in terms of accuracy and performance.

Unlike video monitoring systems, infrastructure-based LiDAR systems are still in their infancy. Zhao et al. [4] utilized roadside LiDAR to track vehicles and pedestrians. They extract LiDAR trajectories of vehicles and pedestrians in the following steps: (1) background filtering, (2) object clustering, (3) vehicle and pedestrian classification, and (4) object tracking. Wu et al. [5] utilized a similar approach to generate object tracks, followed by the identification of vehicle-pedestrian near-crashes. The work [6] computes operational safety assessment metrics using multisensor data from infrastructure. Their multimodel perception sensor system includes fixed video cameras and temporarily deployed sensors, namely a drone with a video sensor, a LiDAR, and a vehicle equipped with differential GPS. They also performed a measurement uncertainty analysis to evaluate the precision of each sensor, using differential GPS as the ground truth. Another study [20] fuses the output of a low-resolution video camera and a solid-state LiDAR and analyzes vehicle-vehicle conflicts. However, given that both types of their sensors are of low resolution, the detection of smaller objects like pedestrians becomes extremely challenging. Zhou et al. [7] applied deep convolutional neural network (DCNN) models trained on autonomous driving datasets to roadside LiDAR data. However, the performance is suboptimal. As stated in the paper, the average F-1 score for vehicles was only 72.9%.

Surrogate safety measures are quantitative indicators that are used to assess the safety of a transportation system instead of using crash data (which are fortunately rare and thus require data collected over multiple years). Surrogate measures can be collected through various technologies, such as video cameras, radar or GPS [8] and can be predictive of crashes and thereby provide measurements of safety. Measures such as time-to-collision (TTC) [21] and post-encroachment time (PET) [22] are two of the most commonly used indicators, as they are based on the temporal proximity of road users. Lower values of TTC and PET indicate higher collision risks. Thresholds of TTC or PET are typically determined based on the perception-reaction time of road users. However, it's worth noting that there is still no consensus on what constitutes a safety-critical event or a near-miss, despite proposals of a hierarchy of traffic events varying in severity. Other measures, such as those based on spatial proximity or acceleration-deceleration patterns of vehicles, have also been proposed [23], [24]. A very comprehensive synthesis of the literature on surrogate safety measures was recently provided by Arun et al. [25]. Surrogate safety measures primarily capture the possible interactions, or "events," among road users, but not all of these events are equally critical for safety. Therefore, it is crucial to distinguish between safe and critical interactions. In this paper, safety-critical events are referred to as "severe events" [8].

Severe events encompass near-miss incidents and unsafe behavior exhibited by road users. The use of surrogate safety measures presents an opportunity to address site-specific and time-specific safety issues and develop countermeasures. However, processing large volumes of video data is a significant practical challenge in utilizing surrogate measures for safety analysis. This involves determining trajectories, identifying conflicts, and filtering critical unsafe maneuvers for further examination. In the context of signalized intersections, analyzing unsafe maneuvers during specific signal phases is crucial to identify appropriate countermeasures. For example, if unsafe maneuvers occur frequently during a permitted left-turn phase, implementing a protected left-turn phase may be necessary. Additionally, conflicts between right-turning vehicles and pedestrians could suggest separating signal phases for these movements, such as disallowing right turns on red or introducing a leading pedestrian phase.

V. Conclusions

We have developed an integrated video and LiDAR analytics system with the aim of improving intersection safety. Advanced deep-learning techniques were employed for both video and LiDAR detection and tracking, which were then mapped to the same Google Maps space. The system performs trajectory clustering, conflict event detection, and severe event playback. Additionally, fusion was performed for both vehicle-to-vehicle and vehicle-to-pedestrian severe conflict events. The overall system serves as an effective tool to help traffic engineers identify potential risks at intersections.

Experimental results for 100+ hours of video and LiDAR data collected concomitantly on the same intersection for the

same week. Video provides rich appearance information that helps people better understand the severity of an event. In addition, the processing speed for video is faster. However, video struggles with night vision and pedestrian detection. LiDAR, on the other hand, is capable of detecting road users with higher precision, performs well during nighttime, and is excellent for pedestrian detection. However, tall objects can cast a large "shadow" in LiDAR point clouds, leading to severe occlusion. Moreover, it is difficult for people to comprehend the scene; for instance, pedestrian crosswalks are not visible. When doing LiDAR annotation, it is beneficial to use video as a reference to classify smaller objects. For example, pedestrians and cyclists are difficult to distinguish in sparse regions. Additionally, some traffic posts and road barriers may also "appear" to be pedestrians in LiDAR point clouds. These results indicate that the two sensors are complementary, and both contribute to a more accurate and reliable system.

In this work, we currently use a union of the severe events found by two modalities. In the future, we plan to intelligently combine the two modalities by automatically choosing the better of the two based on location of the intersection (e.g., areas where LiDAR is more accurate than video and vice versa), traffic conditions (e.g., when video accuracy is reduced due to occlusions), and lighting and weather conditions.

In the past, we have used a fisheye video-based system to study performance and safety for multiple signal timing plans and allow the user to choose appropriate tradeoffs [26]. We plan to extend this work for the hybrid system described above.

ACKNOWLEDGMENTS

This work is supported by NSF CNS 1922782, by the Florida Dept. of Transportation (FDOT), and FDOT District 5. The opinions, findings and conclusions expressed in this publication are those of the author(s) and not necessarily those of the Florida Department of Transportation or the National Science Foundation.

REFERENCES

[1] L. Blincoe, T. R. Miller, J.-S. Wang, D. Swedler, T. Coughlin, B. Lawrence, F. Guo, S. Klauer, and T. Dingus, "The economic and societal impact of motor vehicle crashes, 2019," NHTSA: Washington, DC, USA, Tech. Rep., 2022.

[2] S. R. E. Datondji, Y. Dupuis, P. Subirats, and P. Vasseur, "A survey of vision-based traffic monitoring of road intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 10, pp. 2681–2698, 2016.

[3] M. S. Shirazi and B. T. Morris, "Looking at intersections: A survey of intersection monitoring, behavior, and safety analysis of recent studies," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 1, pp. 4–24, 2016.

[4] J. Zhao, H. Xu, H. Liu, J. Wu, Y. Zheng, and D. Wu, "Detection and tracking of pedestrians and vehicles using roadside lidar sensors," *Transportation Research Part C: Emerging Technologies*, vol. 100, pp. 68–87, 2019.

[5] J. Wu, H. Xu, Y. Zheng, and Z. Tian, "A novel method of vehicle-pedestrian near-crash identification with roadside lidar data," *Accident Analysis & Prevention*, vol. 121, pp. 238–249, 2018.

[6] N. Altekar, S. Como, D. Lu, J. Wishart, D. Bruyere, F. Saleem, and K. L. Head, "Infrastructure-based sensor data capture systems for measurement of operational safety assessment (osa) metrics," *SAE International Journal of Advances and Current Practices in Mobility*, vol. 3, no. 2021-01-0175, pp. 1933–1944, 2021.

[7] S. Zhou, H. Xu, G. Zhang, T. Ma, and Y. Yang, "Leveraging deep convolutional neural networks pre-trained on autonomous driving data for vehicle detection from roadside lidar data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 22 367–22 377, 2022.

[8] T. Banerjee, K. Chen, A. Almaraz, R. Sengupta, Y. Karnati, B. Grame, E. Posadas, S. Poddar, R. Schenck, J. Dilmore, S. Srinivasan, A. Rangarajan, and S. Ranka, "A modern intersection data analytics system for pedestrian and vehicular safety," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 2022, pp. 3117–3124.

[9] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[10] N. Wojke, A. Bewley, and D. Paulus, "Simple online and real-time tracking with a deep association metric," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 3645–3649.

[11] X. Huang, T. Banerjee, K. Chen, N. V. S. Varanasi, A. Rangarajan, and S. Ranka, "Machine learning-based video processing for real-time near-miss detection." in *VEHITS*, 2020, pp. 169–179.

[12] T. Yin, X. Zhou, and P. Krahenbuhl, "Center-based 3d object detection and tracking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 784–11 793.

[13] Z. Pang, Z. Li, and N. Wang, "Simpletrack: Understanding and rethinking 3d multi-object tracking," *arXiv preprint arXiv:2111.09621*, 2021.

[14] A. Wu, P. He, X. Li, K. Chen, S. Ranka, and A. Rangarajan, "An efficient semi-automated scheme for infrastructure lidar annotation," *arXiv preprint arXiv:2301.10732*, 2023.

[15] O. D. Team, "Openpcdet: An open-source toolbox for 3d object detection from point clouds," https://github.com/open-mmlab/OpenPCDet, 2020.

[16] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, no. 4, pp. 325–340, 1987.

[17] S.-C. S. Cheung and C. Kamath, "Robust background subtraction with foreground validation for urban traffic video," *EURASIP J. Adv. Signal Process*, vol. 2005, p. 2330–2340, jan 2005. [Online]. Available: https://doi.org/10.1155/ASP.2005.2330

[18] X. Huang, P. He, A. Rangarajan, and S. Ranka, "Intelligent intersection: Two-stream convolutional networks for real-time near-accident detection in traffic video," *ACM Trans. Spatial Algorithms Syst.*, vol. 6, no. 2, jan 2020. [Online]. Available: https://doi.org/10.1145/3373647

[19] B. Qi, W. Zhao, H. Zhang, Z. Jin, X. Wang, and T. Runge, "Automated traffic volume analytics at road intersections using computer vision techniques," in *2019 5th International Conference on Transportation Information and Safety (ICTIS)*, 2019, pp. 161–169.

[20] A. M. Anisha, M. Abdel-Aty, A. Abdelraouf, Z. Islam, and O. Zheng, "Automated vehicle-to-vehicle conflict analysis at signalized intersections by camera and lidar sensor fusion," *Transportation Research Record*, p. 03611981221128806, 2022.

[21] J. C. Hayward, "Near-miss determination through use of a scale of danger," *Highway Research Record*, vol. Issue 384, pp. 24–34, 1972.

[22] B. L. Allen, B. T. Shin, and P. J. Cooper, "Analysis of traffic conflicts and collisions," *Transportation Research Record*, vol. Issue 667, pp. 67–74, 1978.

[23] S. S. Mahmud, L. Ferreira, M. S. Hoque, and A. Tavassoli, "Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs," *IATSS Research*, vol. 41, no. 4, pp. 153–163, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0386111217300286

[24] X. Shi, Y. Wong, M. Li, and C. Chai, "Key risk indicators for accident assessment conditioned on pre-crash vehicle trajectory," *Accident Analysis and Prevention*, vol. 117, pp. 346–356, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000145751830191X

[25] A. Arun, M. M. Haque, A. Bhaskar, S. Washington, and T. Sayed, "A systematic mapping review of surrogate safety assessment using traffic conflict techniques," *Accident Analysis and Prevention*, vol. 153, p. 106016, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001457521000476

[26] A. Mishra, K. Chen, S. Poddar, E. Posadas, A. Rangarajan, and S. Ranka, "Using video analytics to improve traffic intersection safety and performance," *Vehicles*, vol. 4, no. 4, p. 1288–1313, 11 2022. [Online]. Available: http://dx.doi.org/10.3390/vehicles4040068